Technical Report 1195

Predictor Development and Pilot Testing of a Prototype Selection Instrument for Army Flight Training

Kenneth T. Bruskiewicz
Personnel Decisions Research Institutes, Inc.

Lawrence C. Katz
U.S. Army Research Institute

Janis Houston and Cheryl Paullin
Personnel Decisions Research Institutes, Inc.

Gavan O'Shea
American Institutes for Research

Diane Damos
Damos Aviation Services

February 2007

20070319412

United States Army Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the Department of the Army Deputy Chief of Staff, G1

Authorized and approved for distribution:

Scott E. Graham

SCOTT E. GRAHAM
Acting Technical Director

Michelle Sams

MICHELLE SAMS
Acting Director

Research accomplished under contract for the Department of the Army

Personnel Decisions Research Institute, Inc.

Technical review by

David M. Johnson, US Army Research Institute
Tonia Heffner, US Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) February 2007	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) July 2005- July 2006
4. TITLE AND SUBTITLE Predictor Development and Pilot Testing of a Prototype Selection Instrument for Army Flight Training		5a. CONTRACT OR GRANT NUMBER DASW01-03-D-0016
		5b. PROGRAM ELEMENT NUMBER 633007
6. AUTHOR(S) Kenneth T. Bruskiewicz (Personnel Decisions Research Institutes, Inc.); Lawrence C. Katz (U.S. Army Research Institute for the Behavioral and Social Sciences); Janis Houston, Cheryl Paullin (Personnel Decisions Research Institutes, Inc.); Gavan O'Shea (American Institutes for Research); Diane Damos (Damos Aviation Services)		5c. PROJECT NUMBER A792
		5d. TASK NUMBER
		5e. WORK UNIT NUMBER 308
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Personnel Decisions Research Institutes, Inc. 650 Third Ave. South Suite 1350 Minneapolis, Minnesota 55402		8. PERFORMING ORGANIZATION REPORT NUMBER Technical Report No. 522
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral & Social Sciences ATTN: DAPE-ARI-IR 2511 Jefferson Davis Highway Arlington, VA 22202-3926		10. MONITOR ACRONYM ARI-RWARU
		11. MONITOR REPORT NUMBER Technical Report 1195
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		
13. SUPPLEMENTARY NOTES Contracting Officer's Representative and Subject Matter POC: Lawrence Katz		

14. ABSTRACT *(Maximum 200 words)*:

As part of a project by the U.S. Army Research Institute for the Behavioral and Social Sciences Rotary Wing Aviation Research Unit (ARI RWARU) to produce a selection instrument for Army flight training, several viable, existing predictor measures were identified and several new predictors were developed. The resulting prototype battery was pilot tested with 80 aviator candidates prior to beginning flight school, who provided performance data and subjective feedback. This pilot test resulted in revisions and decisions as to the predictors to be included in the prototype battery for preliminary validation.

15. SUBJECT TERMS

Aviator selection; military; personality; motivation; psychomotor; cognitive; predictor measures

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	44	Ellen Kinzer Technical Publication Specialist (703) 602-8047

Technical Report 1195

Predictor Development and Pilot Testing of a Prototype Selection Instrument for Army Flight Training

Kenneth T. Bruskiewicz
Personnel Decisions Research Institutes, Inc.

Lawrence C. Katz
U.S. Army Research Institute

Janis Houston and Cheryl Paullin
Personnel Decisions Research Institutes, Inc.

Gavan O'Shea
American Institutes for Research

Diane Damos
Damos Aviation Services

Rotary-Wing Aviation Research Unit
William R. Howse, Chief

U.S.Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

February 2007

Army Project Number	**Personnel Performance**
633007A792	**and Training**

Approved for public release; distribution is unlimited

PREDICTOR DEVELOPMENT AND PILOT TESTING OF A PROTOTYPE SELECTION INSTRUMENT FOR ARMY FLIGHT TRAINING

EXECUTIVE SUMMARY

Research Requirement:

In June 2004, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) was tasked with conducting the research and development for a new Selection Instrument for Flight Training (SIFT). The Army's stated objectives were: 1. Develop a computer-based and web-administered selection instrument for Army flight training with emphasis upon aptitudes for Future Force aviator performance within the Future Combat Systems environment; 2. Develop an aviator selection instrument that corrects or minimizes risks associated with several deficiencies identified in the current selection instrument – the Alternate Flight Aptitude Selection Test (AFAST); 3. Develop the selection instrument so that the Army will be able to rapidly assess its current performance as a predictor, revise the instrument when necessary, and adapt its application to selection for related occupational categories such as Unmanned Aerial Vehicle Operators and Special Operations Aviators; and 4. Maximize utilization (by inclusion or adaptation) of existing tests as may be found in use or under development within the Department of Defense. The first task was to review the relevant selection literature to collect information that could be used to produce a rational recommendation for a specific selection and testing strategy for Army aviation (Paullin, Katz, Bruskiewicz, Houston, & Damos, 2006). The second task in this project was to conduct a job analysis for Army aviators to collect information regarding the personal attributes that should be required of flight school candidates (Kubisiak & Katz, 2006). The recommended selection strategy that resulted from these two tasks outlined several viable, existing predictor measures, as well as several new predictors that could be developed.

Procedure:

The researchers determined which of the existing predictors should be included in the prototype aviator selection battery (for subsequent validity testing) and which new predictors should be developed. Existing measures included a general-intelligence-based cognitive test, a measure of motivation, and some existing scales for a newly developed biographical data inventory. New predictors that could be developed within the time frame and resources of the current contract included measures of task prioritization, perceptual speed and accuracy, motivation to become an Army aviator, and several personality traits. The resulting prototype battery was pilot tested with 80 Warrant Officers and Commissioned Officers prior to beginning flight school, who provided test performance data and subjective feedback. This pilot test resulted in revisions to the subtests, as well as decisions as to the predictors to be included in the prototype battery for preliminary validation.

Findings:

Research clearly suggests that cognitive aptitude, or general intelligence (g), will be an important predictor of aviator performance. This pilot test of candidate predictors indicated that the following measures may efficiently and effectively add incremental validity beyond a measure of general intelligence: 1) The Army Aviation Biodata Inventory (AABio), a forced-choice biographical inventory measuring such attributes as adaptability, stress tolerance, and reasonable risk-taking; 2) The Army Aviation Measure of Individual Motivation (AAMIM), including measures of adjustment, agreeableness, dependability, leadership, physical condition, and work orientation; 3) A pared-down version of the Army Aviation Information Test (AAInfo), with multiple-choice questions designed to measure eight content areas (e.g., basic flight rules/knowledge); 4) Two out of three measures of perceptual speed and accuracy (Hidden Figures and Simple Drawings); and 5) An expanded (15-trial) measure of task prioritization, called "the Popcorn Test."

Utilization and Dissemination of Findings:

The recommended selection strategy that emerged from the job analysis and review of selection literature is a two-stage testing process. The first stage involves measuring cognitive and personality/motivational traits important for the aviator job. These tests do not require any non-standard computer peripherals and can be administered via the Internet in virtually any location with access to a desktop computer, the Internet, and a test proctor. This pilot research identified the existing measures and assessed the newly developed measures that appear to offer the best potential for effectively and efficiently predicting Army aviator training performance. The measures were revised, based upon the data generated by this research, and the next step in the development of SIFT will be to evaluate the predictive validity of the prototype test battery.

The second stage of the test battery may include performance-based measures of psychomotor and information processing skills. These tests do require non-standard computer peripherals and may better serve the needs of Army aviation as classification instruments, for tracking selected aviators into one of the four mission platforms. In addition, the systematic test development process described in this report will be used to develop a selection instrument for Army Unmanned Aviation Systems operators.

The results of this research will be disseminated to the developers of the individual tests assessed in the interest of contributing to the literature addressing the use of these measures. In addition, the findings will be communicated to interested parties in the U.S. Army Proponency Office to facilitate their planning for fielding of the prototype test battery that will constitute the final SIFT product, following preliminary validation. As an integral part of the aviator selection test battery development process, the research described herein will be presented to professional organizations interested in military selection such as the Department of Defense Human Factors Engineering Technical Advisory Group and the Army Science Conference.

PREDICTOR DEVELOPMENT AND PILOT TESTING OF A PROTOTYPE SELECTION INSTRUMENT FOR ARMY FLIGHT TRAINING

CONTENTS

APPENDICES

LIST OF TABLES

PREDICTOR DEVELOPMENT AND PILOT TESTING OF A PROTOTYPE SELECTION INSTRUMENT FOR ARMY FLIGHT TRAINING

Introduction

In June 2004, the US Army Research Institute for the Behavioral and Social Sciences (ARI) was tasked with conducting the research and development for a new Selection Instrument for Flight Training (SIFT). The Army's stated objectives were: 1) Develop a computer-based and web-administered selection instrument for Army flight training with emphasis upon aptitudes for current aviator performance; 2) Develop an aviator selection instrument that corrects or minimizes risks associated with several deficiencies identified in the current selection instrument – the Alternate Flight Aptitude Selection Test (AFAST); 3) Develop the selection instrument so that the Army will be able to rapidly assess its performance as a predictor, revise the instrument when necessary, and adapt its application to selection for related occupational categories such as Unmanned Aviation System Operators and Special Operations Aviators; and, 4) Maximize utilization (by inclusion or adaptation) of existing tests as may be found in use or under development within the Department of Defense. To provide assistance to meet these objectives, ARI awarded a contract to Personnel Decisions Research Institutes (PDRI).

The project was divided into several tasks. The first task was to review the relevant selection literature to collect information that could be used to produce a rational recommendation for a specific selection and testing strategy for Army aviation (Paullin, Katz, Bruskiewicz, Houston, & Damos, 2006). The second task in this project was to conduct a job analysis for Army aviators to collect information regarding the personal attributes that should be required of flight school candidates (Kubisiak & Katz, 2006). The recommended selection strategy that resulted from these two tasks outlined several viable, existing predictor measures, as well as several new predictors that could be developed.

The researchers then determined which of the existing predictors should be included in the validation test (based upon coverage of the relevant attributes and practical considerations) and which new predictors should be developed. Some of the new predictors could be developed within the time frame and scope of the current contract, including measures of task prioritization, perceptual speed and accuracy, motivation to become an Army aviator, and several personality traits. Predictors that were not developed under the current contract, but that may be considered for future development, include one or more measures of psychomotor skills and multiple-task performance.

Table 1 provides an overview of the predictor measures included in this research, along with an indication of which ones are new measures and which ones are existing measures. In the following sections, we describe each predictor and, for the new predictors, the process we used to develop and pilot test the items.

Table 1

Overview of Predictor Measures

Test/Inventory	Subtest/Scale	# Items	Existing/New
Army Aviation Cognitive Test (AACog)	Reading Comprehension	27 (25 min time limit)	Existing Navy test
	Math Skills	30 (25 min time limit)	Existing Navy test
	Mechanical Comprehension	30 (15 min time limit)	Existing Navy test
	Spatial Apperception	25 (10 min time limit)	Existing Navy test
	Aviation & Nautical Information	30 (25 min time limit)	Existing Navy test
Popcorn Test	[none]	10-15 trials (10-15 min time limit)	New test
Perceptual Speed & Accuracy (PSA)	Hidden Figures	30	New test
	Simple Drawings	100	New test
	Panel Displays	40	New test
Army Aviation Information Test (AAInfo)	[none]	50	New Test
Army Aviation Measure of Individual Motivation (AAMIM)	Adjustment	25	Existing Army scale
	Agreeableness	20	Existing Army scale
	Dependability (mostly Non-Delinquency)	22	Existing Army scale
	Leadership	23	Existing Army Scale
	Physical Condition	10	Existing Army scale
	Work Orientation	24	Existing Army scale
	Random Response	4	Existing Army scale
	Lie (Unlikely Virtues)	8	Existing Army scale
	Practice	4	Existing Army scale
	AAMIM Subtotal (# Statements)	***140***	
	AAMIM Subtotal (# Tetrads)	***35***	

Table 1. Overview of Predictor Measures (continued)			
Test/Inventory	**Subtest/Scale**	**# Items**	**Existing/New**
Army Aviation Biodata (AABio)	Adaptability	8	New scale
	Army Aviation Identification	7	New scale
	Attention to Detail	13	New scale
	Attitude Toward Authority	13	Existing Army scale
	Cognitive Flexibility	10	Existing Army scale (with some new items)
	Decisiveness	9	New scale
	Diplomacy	5	Existing Army scale
	Fitness Motivation	7	Existing Army scale
	Internal Control	10	New scale
	Multi-Tasking	7	New scale
	Peer Leadership	9	Existing Army scale
	Reasonable Risk-Taking	11	New scale
	Risk Tolerance	10	New scale
	Stress Tolerance	16	New scale
	Work Motivation	12	Existing Army scale
	Lie (Unlikely Virtues)	7	Existing Army scale
	AABio Subtotal	***154***	
Data Checks (items appended to end of AABio; not intended for operational use)	Prior Army Aviation Knowledge	2	New items
	Aviation Experience	2	New items
	Computer/Mouse Experience	3	New items
	Video/Flight Simulation Gaming Experience	2	New items
	Test-Taker Reactions	2	New items

Army Aviation Cognitive Test (AACog)

The published selection research indicates that cognitive aptitude, or general intelligence (*g*), is consistently an important predictor of aviator performance. The recommendation that stemmed from the review of existing predictor measures (Paullin, Katz, Bruskiewicz, Houston, & Damos, 2006) was that the Army should use either the cognitive tests from the U.S. Navy's Aviator Selection Test Battery (ASTB) or the USAF's Air Force Officer Qualification Test (AFOQT) as its cognitive aptitude predictor measure. The Army chose the ASTB, in large part

because it is already web-enabled, and was granted permission from the Navy to administer it via the Navy's web-based system (called the Automated Pilot Examination System, or "APEX") during the preliminary validation reserach. The current version of the ASTB includes subtests measuring:

Reading Comprehension: items require examinees to extract meaning from text passages. Each item requires the examinee to determine which of the response options can be inferred from the passage.

Math Skills: items evaluate examinees' arithmetic, algebraic, and geometry knoweldge. The assessments include both equations and word problems. Some items require solving for variables, others are time and distance problems, and some require the estimation of simple probabilities. Skills assessed include basic arithmetic operations, algebraic operations, fractions, roots, exponents, and the calculation of angles, area, and perimeter of geometric shapes.

Mechanical Comprehension: items assess examinees' knowledge of topics that would typically be found in an introductory high school physics course and the application of these topics within a variety of situations. The questions gauge examinees' knowledge of principles related to gases and liquids, and their understanding of the ways in which these properties affect pressure, volume, and velocity. The subtest also includes questions that relate to the components and performance of engines, principles of electricity, gears, weight distribution, and the operation of simple machines, such as pulleys and fulcrums.

Spatial Apperception: items evaluate an examinee's ability to match external and internal views of an aircraft based on visual cues regarding its direction and orientation relative to the ground. Each item consists of a view from inside the cockpit, which the examinee must match to one of five external views. These items capture the ability to visualize the orientation of objects in three-dimensional space.

Aviation & Nautical Information: items assess an examinee's familiarity with aviation history, nautical terminology and procedures, and aviation-related concepts such as aircraft components, aerodynamic principles, and flight rules and regulations.

The pilot test described herein used the operational version that currently is in use by the Navy to select aviators. Thus, the number of items and time limit for each subtest was determined by the Navy.

In return for allowing the Army to administer the ASTB via APEX, the Navy requested that this pilot test include all five of the ASTB cognitive subtests, although some of the items in the Aviation and Nautical Information subtest may demonstrate low face validity for Army aviation examinees. The Navy wished to examine a new data source, and the Army felt it would be worthwhile to examine performance on their Information subtest. Both Navy and Army researchers agreed that the Army would develop its own aviator selection score composite, and that the Army's composite score might not include all of the ASTB subtests.

Task Prioritization (Popcorn Test)

Pilots frequently perform several tasks concurrently. To assess the timesharing skills and abilities needed to perform concurrent tasks, multiple-task tests have been included in military pilot selection batteries since World War II (Melton, 1947). The multiple-task tests used in pilot selection usually are composed of two tasks, and candidates typically are told the relative priorities of the two tasks. Very few multiple-task tests manipulate task priorities during the assessment period.

There is little question that, to be a successful aviator, a pilot must be able to prioritize concurrent tasks correctly. Additionally, he or she must be able to recognize the need to change his or her task prioritization and adjust the prioritization strategy correctly (Roscoe, 1980). However, creating a selection instrument to identify candidates who can prioritize concurrent tasks quickly and correctly and then adjust the priorities in response to changing conditions is far more difficult than creating an instrument that assesses multiple-task skills. To assess task prioritization, investigators need to have a method for determining the optimum task prioritization of the selection instrument. They also must develop metrics that reflect task prioritization accurately. The selection instrument itself must then be sufficiently difficult to measure individual differences in performance, and these differences would ideally be unaffected by prior experiences such as education, flight experience, exposure to computer games, etc. Preferably, the instrument would be based on a theory-driven model of performance that permits a candidate's performance to be compared to a normative value as well as to that of other candidates. Finally, to determine if a candidate can recognize a changing environment and change his/her priorities accordingly, investigators need to have some method for changing the task priorities.

The test characteristics described above are sufficiently exacting that little work was conducted on task prioritization until the mid-1980s. During this period, NASA sponsored a few theoretical studies of task prioritization (Daryanian, 1980; Pattipati, Kleinman, & Ephrath, 1983; Tulga & Sheridan, 1980). These investigators developed a task known as "Popcorn," the name reflecting the appearance of target stimuli "popping" in a horizontal direction across the screen. Later, NASA investigators conducted one project with a more aviation-oriented version of Popcorn (Hart, Battiste, & Lester, 1984). NASA-sponsored research on task prioritization appears to have stopped at this point. The only subsequent work on task prioritization was done by Roscoe and his colleagues (Roscoe, Corl, & LaRoche, 1997). They developed a selection task, WOMBAT, which is currently the only commercially available selection instrument that purports to measure prioritization. The cost of WOMBAT was found to be prohibitive for the purposes of Army aviation selection, and to date, it has not been used in military pilot selection.

It should be noted that none of the NASA-sponsored work was concerned with selection. Consequently, no data are available on the test-retest reliability or predictive validity of Popcorn. Additionally, the relationship between Popcorn and other selection instruments, such as intelligence tests, was never investigated.

This project's subcontractors, Damos Aviation Services (DAS) and the American Institutes for Research (AIR), collaborated to develop and program a Popcorn Test designed to measure the ability to prioritize tasks that are occurring quickly and simultaneously. In this test, blocks of varying size move horizontally across the computer screen from left to right at varying rates of speed. There are five lines upon which blocks may be moving at any given time, with up to three blocks moving on each line at any given time. The test-taker uses a mouse to control an on-screen cursor. The test-taker scores points by "erasing" each block before it reaches the right edge of the stimulus box. Targets are erased by simply holding the cursor on the block continually while the block is gradually eliminated. Larger blocks are worth more points than smaller blocks and faster-moving blocks are worth more points than slower-moving blocks. Popcorn displays nine sizes of targets (all possible combinations of one, two, and three units of height and length, where a unit is equal to 0.3 in.). The candidate only scores points when a target is completely erased. Target length, height, speed, and arrival time are generated by a random number generator that produces a square distribution. Target length and height use the same seed for the random number generator; all other parameters use unique seeds. Block size and speed are multiplied to obtain a final score for each of ten, 90-second trials. Prior to the scored portion of the test, test-takers receive detailed instruction and one 30-second practice trial. There is also a 20-second rest break between successive scored trials.

Because little, if any, work has been conducted on task prioritization in an aviation context since the mid-1980s, the current version of Popcorn is based on the four NASA-sponsored studies described earlier. None of the Popcorn tasks described in these four studies could be used "as is" for two reasons. First, the authors described many of the task characteristics in terms of the computer technology of the time. These descriptions cannot be easily translated to the existing technology. Second, the training and testing times described in these studies were too long. Although some of the authors did not specifically mention the training time, all of the versions appear to have required at least one hour of practice and testing required 50 minutes or more. To avoid these extreme training and testing times, a simplified version of Popcorn, described above, was developed for this research.

This version of Popcorn does not manipulate the experimental conditions across trials; such a version would require a much longer development time and an extended training and testing period. The instructions tell the candidates explicitly that the points that can be earned for each target are equal to the area of the target multiplied by its speed. Telling the candidates how the points are calculated is intended to reduce the amount of time needed to develop a strategy. However, since all of the candidates know how the points are calculated, the individual differences in performance that would have been due to figuring this out are removed.

Performance metrics were a major focus of this developmental effort. Two basic types of metrics were developed. The first, which is called the "composite," is the ratio of the number of points the candidate obtained in a trial divided by the total number of possible points. The total number of possible points is the sum of the points associated with each of the targets presented during the trial. This "composite" score is presented on-screen to the test taker after each trial.

The second performance measure is called "indecision." This measure reflects the number of times the candidate began erasing a target but stopped before the target was

completely erased and did not return to the target. The indecision measure was included because some individuals appear to have poor target selection strategies and move frequently from target to target without erasing many targets.

Indecision currently has two primary problems. First, a candidate can accidentally move his/her cursor through a target while moving to the intended target. Depending on how quickly the candidate moves the cursor, the system may register the transit and increase the indecision counter (thus "penalizing" the candidate). Therefore, for this pilot test the minimum time on target was increased from 100 ms to 250 ms to account for this situation.

Second, in computing the indecision measure, a candidate may be penalized for using an optimal strategy. For example, assume that the candidate is erasing one target and realizes that a target with more points (higher priority) has just been displayed. The candidate switches to the higher priority target and erases it. By the time the candidate finishes erasing the high priority target, the first target has reached the end of the track. Although this strategy may be optimal, the candidate would be penalized for not completing the erasure of the first target. No countermeasure for this problem was instituted for the pilot test, and therefore the indecision measure would have to be considered with caution.

Popcorn Pretest

Pre-testing of Popcorn was conducted to refine the instructions and parameter settings and to determine the optimal length of the trials and the inter-trial breaks. The pretest also explored potential problems with the mouse and the physical layout of the testing station. Following pretest sessions, subjects were questioned about their strategies and the layout of the testing station. Several questions were designed to determine how well they understood the instructions.

The sessions conducted at AIR were held in a computer lab with five terminals, which allowed participants to be tested concurrently. Over three days, 13 sessions were conducted with 24 AIR research assistants (11 male, 13 female) and 20 additional college-aged and high-school aged recruits (19 male, 1 female). AIR employees completed the sessions after work hours and received a $20 gift certificate for their participation. Non-AIR employees were paid $40, and the highest scorer from this group also received a $200 prize. The sessions conducted at PDRI were conducted in a conference room on laptop computers. Thirteen participants participated in the PDRI pretest (9 male, 4 female) and participants were each paid $40. All of the participants were approximately high-school aged.

Following each day's sessions, the data were examined and appropriate adjustments to the Popcorn test parameters (e.g., the targets' presentation rate and speed) were made. Analyses of the data from the first group of subjects demonstrated that the task was learned too easily and that the subjects received too much practice before the actual test trials began. The practice was shortened to one, 30-second trial. The parameters were manipulated to make the test more difficult by increasing the speed of the targets and making them appear more frequently. By the end of the pre-testing, the upper cutoff value for the arrival times of the target (maximum

inter-arrival time) was 1.9 seconds. The target speeds were set to range from 0.5 to 2.9 inches/second.

The instructions for Popcorn were revised and refined throughout the testing, but had been finalized by the conclusion of pretest administration. The post-testing questions indicated that many participants had developed a good strategy by the third or fourth trial. Interestingly, several of the participants could specify exactly how the target points were computed but then described using a suboptimal strategy.

One of the important characteristics explored with the pretest data was differential stability. Any dependent measure obtained from an information processing test using short, repeated trials should have obtained differential stability before it can be used as a predictor. Differential stability occurs when: 1) the group mean performance is either constant or increasing in a slow, linear fashion; 2) the trial-to-trial variances are constant; and 3) the rank order of individuals is constant within some pre-specified level of error. Before a task reaches differential stability, the trial intercorrelation matrix typically demonstrates superdiagonal form, that is, the inter-trial correlations decrease across a row and down a column.

By the end of the pre-testing sessions, the inter-trial correlation matrix of Popcorn did not demonstrate superdiagonal form. Additionally, mean performance on some of the trials seemed to deviate significantly from an apparent trend and the associated variance was substantially larger than that of neighboring trials. It was decided that these differences could be attributed to between-trial differences in the total number of targets presented and their average speed. Although several trials had been deleted and replaced during the contractor pre-testing, another trial was deleted and replaced with an existing trial for the Army pilot test.

The final Popcorn version included in the pilot test battery included the testing components of the pre-tested lab version, but was designed to run in a browser over the Internet. As such, results were reported to a server, which then stored them in a database. Finally, various optimization methods were employed to minimize the potential problems associated with running the program in an interpreted environment (e.g., Javascript).

Perceptual Speed and Accuracy (PSA)

The recommended selection strategy included measures of perceptual speed and accuracy. Three types of items were developed for the assessment of PSA: Hidden Figures, Simple Drawings, and Panel Displays.

Hidden Figures

The Hidden Figures test measures the extent to which the examinee can distinguish simple shapes or objects that are "hidden" from obvious view by interfering lines and shapes in a more complex object, often referred to as Field Independence or Figure/Ground skill. This construct has been defined as the ability to hold the stimulus shape or object in mind so as to distinguish it from other well-defined perceptual material. Such a test was developed for use during the Army's Project A to develop and validate new selection tests (Russell, Peterson, Rosse, Hatten, McHenry, & Houston, 2001). This construct appears to be relevant for all pilots,

including helicopter pilots, as they look for objects on the ground from various perspectives aloft (Kubisiak & Katz, 2006).

Each of the 30 items in the Hidden Figures test requires the test taker to determine which of five simple figures (presented at the top of the screen) is hidden within a complex pattern. Only one of the five simple figures is included in each complex pattern, and the figure is always right side up and the same size as in the drawings at the top of the screen. This test is scored as the number correct minus a fifth of the number incorrect (i.e., there is a correction for guessing). There is a 6-minute time limit for this test.

Simple Drawings

The Simple Drawings test is the more typical perceptual speed and accuracy measure (Arth, Steuck, Sorrentino, & Burke, 1990). Each item is a set of five simple drawings, mostly non-real-world or unidentifiable objects, where one and only one is not identical to the other four. The test taker must indicate which of the five drawings is unlike the other four. There are 100 items of this type, with a two-minute time limit. Again, there is a correction for guessing in the scoring for this test.

Panel Displays

The third PSA test, Panel Displays, was developed for this project. It is identical in concept to the Simple Drawings, but the objects are actual helicopter gauge displays. In each item, a set of five gauges is presented, with one and only one gauge in a slightly different configuration from the other four, which are identical. A total of 40 items were developed for this test and a time limit of two minutes was set. There is a correction for guessing for this test as well.

Motivation to Become an Aviator

The recommended selection strategy included a measure of motivation and attitudes toward becoming an aviator. Pilot selection researchers have often measured motivation to become an aviator using a knowledge test format, that is, multiple-choice test questions that assess knowledge of aviation topics. For example, the ASTB, AFOQT, and AFAST all include an Information subtest that uses a multiple-choice knowledge test format. The logic is that persons who are more motivated to become an aviator will make an effort to learn about aviation and will thus possess more aviation knowledge than persons with lower motivation levels. Another way to measure motivation and attitudes toward becoming an Army aviator is to use a direct self-report approach.

It was decided to try both an indirect, knowledge-based approach and a direct measurement approach. First, a knowledge test called the Army Aviation Information Test (AAInfo) was developed. Second, self-report items were developed, aimed at directly measuring motivation to become an Army aviator, and this scale was included in the biodata inventory described in a later section. The following section describes the AAInfo test.

Army Aviation Information Test (AAInfo)

As noted, the current Army pilot selection battery, the AFAST, includes a helicopter knowledge subtest. The helicopter knowledge subtest scores of all candidates who graduated from the Army's Initial Entry Rotary Wing (IERW) training course between January 1988 and February 1993 were collected. Analyses of these data were conducted separately for Warrant Officers (N = 1,052) and Commissioned Officers (N = 605). Within each group, scores on the helicopter knowledge subtest ranged from zero to the maximum score of 20. The mean scores were 12.61 (SD = 3.97) and 11.26 (SD = 4.08) for Warrant Officers and Commissioned Officers respectively. These scores are from individuals who were selected into and graduated from flight training, so their average scores are probably higher than the average score across the entire applicant pool.

These data confirmed that, as of 15 years ago, there was a good distribution of scores on the AFAST helicopter knowledge subtest. However, aviation information is more accessible to applicants today than it was 15 years ago. For example, a great deal of relevant information can be located via the Internet, in addition to traditional sources such as libraries, flight schools, and trade journals. As a consequence, there may be a higher overall level of helicopter knowledge among the applicant pool, but there may also be more variability in the amount of knowledge that applicants possess. When developing the AAInfo tests, the AFAST helicopter knowledge test served as a foundation, but all of the items were rewritten (for test security reasons) and the scope of the test was broadened by writing additional items of widely varying degrees of difficulty.

A knowledge test of this type is only an indirect measure of motivation. Army aviation applicants are not required to possess knowledge of aviation topics or flight experience prior to entering aviator training, so differences among them in level of motivation should be reflected in scores on a well-crafted knowledge test --to the extent that motivation is exhibited by learning about the topic in which one is interested. The knowledge test format offers the advantage of being a "non-fakable" measure of motivation. Examinees either know or do not know the answers to the questions; they cannot claim to be more motivated than they really are. In psychometric parlance, a knowledge test can be viewed as a *maximal* measure of motivation, rather than *typical* measure of motivation (Held & Farmer, 2004). General intelligence certainly impacts the extent to which individuals gain and can recall knowledge, so it might be expected that scores on the AAInfo test would correlate with scores on the ASTB subtests, particularly the Aviation and Nautical Information subtest.

Development of the AAInfo test began with the creation of an item-writing plan, as shown below:

1. Number of items: 50 items for tryout, final version likely to be approximately 20-25 items

2. Item type: 5-option multiple choice (matches the number of options in ASTB and AFAST information subtests)

3. Item difficulty: range of difficulties

4. Reference materials: base the items on well-known and widely-available official reference sources, but make certain the information is available in multiple sources, to ensure that performance isn't heavily impacted by which reference materials a person happens to have available to study

5. Starting point: AFAST helicopter knowledge items – revised to make them different from current versions and to develop a non-redundant set of items; about a third to a half of the 50-item target were derived from existing AFAST items

6. Content coverage: The AFAST helicopter knowledge test was content analyzed to assure that the same topic areas were covered; new content areas were also added to broaden the scope of the AAInfo test

Table 2

Content Area Coverage of the AAInfo Test

AAInfo Test Content Area	# items
Major helicopter controls and parts, and their functions (covered in AFAST helicopter knowledge test)	12-14
Basic operation of a helicopter (covered in AFAST helicopter knowledge test)	10-12
Physical forces impacting helicopter flight (covered in AFAST helicopter knowledge test)	5-7
Meteorological conditions impacting helicopter flight (covered in AFAST helicopter knowledge test)	3-5
Basic flight rules & knowledge (new content area)	4-5
Specific types of Army helicopters, and major distinctions among them (new content area)	2-4
Information specific to Army aviation, but not about helicopters (new content area)	2-4
Work conditions faced by helicopter pilots (new content area)	2-5

Based on the findings of the job analysis, a draft pool of items was written according to the test plan shown in Table 2. For each item, including those modeled after the AFAST helicopter knowledge test, at least one reference source that covers the knowledge tapped by that item was consulted. Most of the items were based on documents published by the Federal Aviation Administration (FAA). A few were based on information published by the U.S. Army. The draft items were reviewed by subject matter experts, and their revisions were incorporated prior to the pilot test.

Army Aviation Measure of Individual Motivation (AAMIM) and Army Aviation Biodata Inventory (AABio)

In the report for the first phase of this project, Paullin, Bruskiewicz, Houston, and Damos (2005), identified four different personality inventories as viable candidates for inclusion in the Army aviator test battery:

1. Test of Adaptable Personality (TAP; Kilcullen, 2004),
2. Assessment of Individual Motivation (AIM; White &Young, 1998),
3. Self Description Inventory (SDI; Christal, 1975),
4. Armstrong Laboratory Aviation Personality Scale (ALAPS; Retzlaff, King, McGlohn, & Callister, 1996).

To administer all four inventories during the preliminary validation research would be too time-consuming, and too fatiguing for test-takers, so a subset of these was selected. The ALAPS scoring key was published in a publicly-accessible technical report, so this inventory was eliminated from further consideration. For the remaining three inventories, the extent to which the inventories covered the personality constructs most important for pilot performance was examined. The Big 5 personality taxonomy, consisting of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness was used as an organizing principle, because it demonstrates the most direct relevance to the job of aviator in the Army today (Grice & Katz, 2006). The personality constructs identified in other research as important for pilots were roughly mapped onto the Big 5 taxonomy, as shown in Table 3. Subject matter experts reached a consensus judgment about which scales measure, to at least some degree, the constructs identified as important for the pilot job. The consensus judgments are also reflected in Table 3. Some of the personality constructs did not fit neatly into the Big 5 taxonomy, so they were placed in a separate category.

Based on the mapping displayed in Table 3, the TAP [for the purposes of this research, disguised by calling it the Army Aviation Biodata Inventory (AABio)] and the AIM [disguised by calling it the Army Aviation Measure of Individual Motivation (AAMIM)] were judged to be the most reasonable measures to administer in the preliminary validation study. They were to be supplemented by new scales that would cover constructs not measured (well) by any of the existing inventories but identified as important to pilot performance in the first phase of this project. The TAP and the AIM were chosen over the SDI because they included scales that were more narrowly focused than the broader SDI scales, with the added advantage that they were developed by U.S. Army researchers and thus were the most easily accessible of all four possible inventories. There is some redundancy in content coverage across the TAP and the AIM, but the two inventories use very different item formats. The TAP consists of biodata items accompanied by Likert-type response scales (e.g., "Very often" to "Never"). The AIM consists of personality statements and uses a forced-choice item format. Each is described in more detail below.

AAMIM Scales. Existing Army scales adapted from the AIM were used to measure each of the following constructs:

1. Adjustment: psychological health versus maladjustment and dysfunction
2. Agreeableness: pleasantness and sociability
3. Dependability: characteristic of being reliable and responsible for one's actions
4. Leadership: ability to command and foster followership
5. Physical Condition: motivation to maximize and maintain physical fitness
6. Work Orientation: employment-related goal-directedness

Table 3

Consensus Judgments of which Measures Tap Constructs Important to Pilot Performance

Construct	TAP	AIM	SDI
Extraversion			Extraversion
Assertiveness/Dominance	Peer Leadership	Dominance	
Achievement Orientation	Work /Fitness Motivation	Work Orientation	
Conscientiousness	Respect for/Hostility Toward Authority (one aspect)	Dependability (scale is broader than Dependability)	Conscientiousness
Responsibility	Work Motivation (to some degree)		
Dependability			
Integrity			
Self-Discipline			
Emotional Stability		Adjustment	Neuroticism
Emotional Stability			
Stress Tolerance			
Agreeableness	Team Orientation	Agreeableness	Agreeableness
Agreeableness/Friendliness	Diplomacy		
Cooperativeness			
Interpersonal Tolerance			
Openness			Openness
Cognitive Flexibility	Cognitive Flexibility		
Non-Big Five Constructs			
Adaptability			
Reasonable Risk-Taking			
Internal Locus of Control			
Self-Confidence/Self-Esteem		Dominance (to some degree) Work Orientation (to some degree)	
Attention to Detail			
Motivation to Become an Army Aviator			

AABio Scales. Existing Army scales adapted from the TAP were used to measure each of the following constructs:

1. Attitude Toward Authority: characteristic relationship with those in positions of power
2. Cognitive Flexibility: ability to change plans or mental models to fit changed circumstances
3. Diplomacy: ability to conduct activities with another with tact in order to bring about a good working relationship
4. Fitness Motivation: motivation to maximize and maintain physical fitness
5. Peer Leadership: ability to command and foster followership among peers
6. Work Motivation: employment-related goal-directedness

When writing new AABio scales, the biodata style of the TAP was emulated so that the items could easily be inserted into that inventory. New AABio scales were written to measure each of the following constructs:

1. Adaptability: ability to change or be changed to fit changed circumstances
2. Army Aviation Identification: belief that one's values, attitudes, and skills are a good fit for the aviator job
3. Attention to Detail: ability to focus on less salient aspects of tasks or the environment
4. Decisiveness: certainty or resoluteness of purpose
5. Internal Locus of Control: perceiving oneself as being in control of one's experiences
6. Multi-Tasking: concurrent operation of two or more processes
7. Reasonable Risk Taking: willingness to expose oneself to potential loss or damage when potential harm is outweighed by potential benefits
8. Risk Tolerance: capacity to endure potential loss or damage when it is inherent to the situation
9. Stress Tolerance: capacity to endure a state of mental or emotional strain or suspense

Risk taking for Army aviators is not the same thing as risk taking for most other jobs. To gain a better understanding of this construct in an aviation context, a focus group was conducted with several highly experienced Army aviators and trainers who explained that Army aviators must be able to accept significant risks to their well-being as a natural part of their job. Often there is no choice about whether or not to take a risk. The only choice is how to act in a manner that effectively accomplishes the mission while minimizing, to the extent possible, risk to the pilot and crew. In an aviation context, "bad" risk-taking behavior generally involves acting impulsively without thinking about the consequences, for example, by showing off. "Bad" risk-taking behavior can be annoying or problematic in any job but it can have very costly and lethal consequences in an Army aviation setting.

Based on the information gained in this focus group, two different scales were written–one focusing on reasonable risk taking tendencies and one focusing on the willingness to tolerate or accept unavoidable risks. Additionally, attempts were made to write the items in such a way that a high score on one scale did not automatically lead to a low score on the other. Empirical evidence will clearly be required to gauge the extent to which this challenge was successfully met.

An earlier technical report for this project (Paullin, Katz, Bruskiewicz, Houston, & Damos, 2006) contains information about the development of the TAP and the AIM. It is worth noting here, however, that the AIM uses a forced choice format. Items are presented in sets of four (called tetrads). Within each tetrad, test-takers must pick the one statement that is most like them and, among the remaining choices, the one statement that is least like them. Within each tetrad, each statement is scored on a different scale. Unlike some forced-choice inventories, every statement is not paired with every possible other statement. This means that it is possible for a test-taker to respond to all of the tetrads in the entire inventory, and never choose an item from a particular scale as most or least like him or her. The original ARI test developers created a scoring system that takes into account this possibility, in a way that ensures all test-takers receive a score on all scales. Those scoring recommendations were used when scoring the AAMIM for the current research effort.

Both the AABio and AAMIM are scored rationally as opposed to empirically. This means that each item is scored such that a higher score indicates a higher (more desirable) level of the personality construct in question. The AAMIM inventory includes items designed to detect random responding (Random Response Scale). Both inventories include items designed to detect extreme exaggeration of positive traits (Lie or Unlikely Virtues Scale) and methods have been designed to correct scale scores for exaggerated responding.

Development of a Web-based Testing Platform

Web-enabled versions of the tests described in the previous sections were developed using Edoceon, an Internet-based data collection system created by AIR. Essentially, Edoceon is a flexible tool that allows end users to build surveys and tests using a variety of different question types. The system consists of three inter-related modules:

Security Module – This feature allows the Edoceon administrator to define users according to their specific roles in the testing process (e.g., administrator, designer, test taker) and to specify the tests that different types of users are able to access.

Design Module – The Design Module enables users to create specific tests. Specifically, users can: a) customize Edoceon's color and layout, b) build item parameters and response formats (e.g., question types, validity constraints), and c) denote skip patterns.

Reporting Module – This module creates a flat file of all raw responses entered by a particular test taker. In addition, the reporting module includes a codebook that specifies the valid range of responses for each question type.

Pilot Testing the New Predictor Measures

Pilot Test Sample

Eighty (80) Army officers participated in the pilot test. Because the goal was to have the pilot test sample conform as closely as possible to the preliminary validation sample, the participants were tested on either the Monday or Tuesday before they started flight school. Thirty-six (36) participants were drawn from the IERW 05-17 class, while the remaining 44 were members of the IERW 05-18 class. The sample consisted of approximately equal numbers of Warrant Officers and Commissioned Officers.

Pilot Test Procedures

Testing conditions were similar to those expected to occur operationally, including a proctored setting, and with all items presented on computers via a website. The pilot test sessions were held in an Army classroom containing 24 computer stations and an instructor computer. Four separate sessions were required to accommodate the 80 participants. Sample sizes for these four sessions were: Session 1 ($n = 20$); Session 2 ($n = 16$); Session 3 ($n = 20$), and; Session 4 ($n = 24$). Detailed specifications appear in Appendix A.

The participants reported to the classroom for the test sessions which were proctored by project personnel. Once all of the participants had arrived, the participants were briefed regarding the research goals and purpose, the testing procedures, and the computer equipment. After participants had read and signed the informed consent form, the proctors distributed User IDs and passwords that allowed them to access the online tests. Once everyone had successfully used this information to log into the test platform, they were told to read the instructions for each test carefully, to try their best, and to raise their hands if they had any questions or computer problems.

The test administration order was designed so that the timed tests occurred first. Following a 15-minute break, participants completed the two un-timed tests at their own pace. During the testing session, the proctors periodically circulated throughout the room to observe the testing procedures, answer questions, and ensure that the online tests were working properly. The orientation and login process took approximately 20 minutes.

Table 4 presents the test administration order that was used for all pilot test sessions. Note that the time limits associated with each test in Table 4 do not include time required to read instructions and complete sample items. Thus, most participants completed the first five tests in 90 minutes. Though the time required by participants to complete the final two self-paced tests following the break varied considerably, nearly all completed the entire battery in three hours.

Table 4

Administration Order for Pilot Test Sessions

Test Name	Time Limit
1. AAInfo	35 min.
2. Popcorn	90 sec. per trial across ten trials
3. PSA: *Simple Drawings*	2 min.
4. PSA: *Panel Displays*	2 min.
5. PSA: *Hidden Figures*	6 min.
15 Minute Break	
6. AAMIM	Un-timed
7. AABio	Un-timed

Participants were told to bring the form containing their User ID and password to one of the proctors once they had completed the final test. When this occurred, the proctors examined the participant's computer screen to verify that they had finished, presented them with a debriefing form, and asked for any impromptu feedback they wished to offer.

Popcorn Debriefing Exercise

Given the relatively novel nature of the Popcorn Test, it was important to verify that participants fully understood the associated instructions. To this end, all participants completed the short questionnaire presented in Appendix B immediately after finishing the Popcorn Test. Of primary interest was whether participants understood that target speed and size jointly determined the number of points earned within a given trial.

Of the 36 participants from the first two pilot test sessions, 18 (50%) correctly perceived that consistently erasing large, fast moving targets garnered the highest score. Nearly all of the remaining participants tended to focus exclusively on either size or speed: most reported that erasing large targets, regardless of their speed, would lead to the highest score (n = 10), though several (n = 4) prioritized fast targets without considering size. All but three participants reported developing a strategy as they completed the Popcorn trials. When asked to describe these strategies, they generally provided more detailed elaborations of their answers to the first question on the debriefing form.

To create a more uniform frame of reference associated with the Popcorn scoring rules during the final two testing sessions, as participants were reading the Popcorn instructions to themselves, the proctor verbally emphasized that "Both the size of the block and the speed of the block impact the score you will receive – large, fast moving blocks are worth the most points."

This verbal adjunct to the written instructions seemed to have a positive impact, as a clear majority of participants in the third and fourth testing sessions correctly interpreted the

multiplicative nature (size x speed) of the Popcorn scoring rubric (n = 37 out of 44, or 84 %). Moreover, five of the remaining seven participants selected one or more options in addition to the correct answer to the first question on the Popcorn debrief form. Responses to the strategy-focused questions (#2 and #3) were very similar to those observed during the first two testing sessions: all but four participants reported generating a strategy, and their descriptions of these strategies largely conformed to their responses to Question #1.

On-Screen Presentation of Popcorn Scores

Recall that the score for each trial is presented on-screen to the test taker following each trial. Originally, the score was presented as a percentage (e.g., 62.58%). In general, most participants appeared to score between 40% and 65% across trials. During the first and second testing sessions, it was clear that receiving such scores caused a fair amount of anxiety and/or annoyance among participants; statements such as "My scores are so low!" and "I'm failing this test!" were not uncommon. In hindsight, these reactions were understandable, given that a score of 70% generally connotes minimally acceptable performance in the Army and in most U.S. educational contexts. To alleviate this potential source of anxiety, the score reporting was modified for the third and fourth testing sessions by removing the percent sign and the decimal point (e.g., 6258 rather than 62.58%). This change reduced score-related discussions in the final two testing sessions and seemed to reduce the anxiety caused by the percentage-based format.

Pilot Test Analyses and Results

As a first step in analyzing the pilot test data, screens for random responding were conducted for the AABio and AAMIM, the Popcorn Test, and the PSA measures. The AAMIM contains a random response item, and the AABio has an Unlikely Virtues scale. Both of these were investigated, and four individuals were removed from the data set because they responded incorrectly to one or the other. For the Popcorn Test, two individuals' scores whose average percent correct on the combined 8th, 9th, and 10th trials was less than 30 percent were removed. Finally, for the PSA measures, the scores were examined for drastic outliers (who may not have been attending, or may have been intentionally answering randomly) but no indicators were found that required removing any participants. Thus, analyses were conducted using a sample of 76 individuals for the AABio and AAMIM measures, 78 individuals for the Popcorn Test, and 80 individuals for the PSA and AAInfo measures. The results are reported below, separately by measure.

Army Aviation Biodata Inventory

The AABio measure included 154 multiple-choice items with five response choices. All AABio items were scored on a continuum from one to five, such that responses representing a more desirable level of the target attribute received a higher score value. Scale scores were created by averaging the scores of the items in each scale. Means, standard deviations, and internal consistency reliabilities (coefficient alpha) were calculated for each scale. Item-total correlations were also calculated within each scale. Items that had item-total correlations of less than .20 were examined to determine whether they needed to be modified or dropped. Finally, scale intercorrelations were calculated.

Table 5 shows the number of items on each scale of the AABio, and the mean, standard deviation, and reliability of the scale scores. In general, the scale scores demonstrated internal consistency reliabilities ranging from .64 to .80. The lowest reliabilities were found for Multi-Tasking, Work Motivation, and Adaptability. The highest reliabilities were found for Attitude Toward Authority, Peer Leadership, and Diplomacy. The scale means ranged from 3.08 (Multi-Tasking) to 4.00 (Internal Control). Thus, not unexpectedly, there was significant negative skew in the scale scores. There were 21 items that had item-total correlations less than .20 and 5 items that had item-total correlations less than .10. A number of these items were in the Adaptability and Multi-Tasking scales, which had low alphas. All of these 26 items were investigated. Of these items, 12 were revised and 4 were deleted and replaced with entirely new items following testing.

Table 5

Descriptive Statistics of AABio Scales

Scale	# Items	Mean	*SD*	Alpha
Adaptability	8	3.68	0.39	.55
Army Aviation Identification	7	3.99	0.58	.71
Attention to Detail	13	3.54	0.39	.69
Attitude Toward Authority	13	3.53	0.43	.79
Cognitive Flexibility	10	3.61	0.48	.68
Decisiveness	9	3.27	0.40	.64
Diplomacy	5	3.66	0.72	.80
Fitness Motivation	7	3.75	0.55	.69
Internal Control	10	4.00	0.42	.64
Multi-Tasking	7	3.08	0.33	.19
Peer Leadership	9	3.53	0.52	.80
Reasonable Risk-Taking	11	3.51	0.47	.74
Risk Tolerance	10	3.60	0.48	.73
Stress Tolerance	16	3.50	0.38	.75
Work Motivation	12	3.55	0.36	.45

$N = 76$.

Table 6 presents the intercorrelations of the AABio scale scores. The scale intercorrelations ranged from low and non-significant to moderately high. The highest correlations were found between Adaptability and Cognitive Flexibility ($r = .65, p < .01$), Peer Leadership and Cognitive Flexibility ($r = .62, p < .01$) and Stress Tolerance and Peer Leadership

($r = .60, p < .01$). Generally speaking, the pattern of intercorrelations was very interpretable. Constructs that should rationally be related to each other were correlated, and vice versa, constructs that should not be related to each other were not significantly correlated (e.g., Stress Tolerance and Reasonable Risk Taking were not significantly correlated with each other). Thus, the AABio scales appear to be relatively homogenous and measuring reasonably independent constructs.

Table 6

Intercorrelations of AABio Scale Scores

	Adaptability	Army Aviation Identification	Attention to Detail	Attitude Toward Authority	Cognitive Flexibility	Decisiveness	Diplomacy	Fitness Motivation	Internal Control	Multi-Tasking	Peer Leadership	Reasonable Risk-Taking	Risk Tolerance	Stress Tolerance	Work Motivation
Army Aviation Identification	.17	—													
Attention to Detail	.29*	.33**	—												
Attitude Toward Authority	.24*	.06	.07	—											
Cognitive Flexibility	.65**	.09	.36**	.14	—										
Decisiveness	.33**	.11	.18	-.08	.30**	—									
Diplomacy	.29*	.10	.14	.21	.32**	.21	—								
Fitness Motivation	.10	.08	.10	.09	.16	.05	.19	—							
Internal Control	.40**	.07	.45**	.12	.48**	.30**	.21	.07	—						
Multi-Tasking	.07	.09	.37**	.06	.08	.12	.23*	-.01	.22	—					
Peer Leadership	.51**	.15	.34**	.18	.62**	.29*	.56**	.28*	.31**	.12	—				
Reasonable Risk-Taking	-.04	.10	.17	.42**	.08	-.31**	.03	-.16	.16	.07	.05	—			
Risk Tolerance	.30**	.10	.22	-.23*	.36**	.57**	.24*	.24*	.33**	-.05	.43**	-.43**	—		
Stress Tolerance	.57**	.18	.53**	.31**	.57**	.39**	.36**	.34**	.56**	.29*	.60**	-.02	.41**	—	
Work Motivation	.29*	.31**	.43**	.40**	.32**	.25*	.14	.26*	.45**	.19	.35**	.30**	.17	.36**	—

$N = 76$. ** Correlation is significant at the $p < .01$ level. * Correlation is significant at the $p < .05$ level.

Army Aviation Measure of Individual Motivation

The AAMIM consisted of 35 items and used a forced choice format. One of the 35 items was used to detect random responding. All items were presented in sets of four (called tetrads). Within a tetrad, test takers must pick the one statement that is most like them and, among the remaining choices, the one statement that is least like them. Within each tetrad, each statement is scored on a different scale. As with the AABio, responses indicating a more desirable level of the target attribute received a higher score value. Specifically, respondents received a score of 2 if they chose a desirable statement as most like them or an undesirable statement as least like them. To calculate the scale scores, the scores associated with the statements included in each scale were averaged.

As with the AABio, we calculated means, standard deviations, and internal consistency reliabilities for the AAMIM scales. We also calculated item-total correlations as well as scale intercorrelations. The descriptive statistics for the AAMIM scores are presented in Table 7. On average, the participants scored slightly higher on the Physical Condition scale than the other scales. All of the scales except for Dependability exhibited internal consistency estimates ranging from .56 to .74. The Dependability scale had an internal consistency estimate of .43.

Table 7

Descriptive Statistics of AAMIM Scales

Scale	**# Items**	**Mean**	***SD***	**Alpha**
Adjustment	25	1.37	.22	.72
Agreeableness	20	1.34	.19	.56
Dependability	22	1.36	.16	.43
Leadership	23	1.31	.22	.74
Physical Condition	10	1.41	.30	.71
Work Orientation	24	1.37	.20	.66

N = 76.

Table 8 presents the intercorrelations of the AAMIM scale scores. The intercorrelations of the AAMIM scales ranged from low and non-significant to moderate. The greatest correlations were found between Agreeableness and both Adjustment and Dependability ($r = .46$ and $r = .45$, respectively, $p < .01$ in both cases). Thus, the AAMIM scales appear to be relatively homogenous and measuring reasonably independent constructs.

Table 8

Intercorrelations of AAMIM Scales

	Adjustment	Agreeableness	Dependability	Leadership	Physical Condition	Work Orientation
Agreeableness	.46**	—				
Dependability	.19	.45**	—			
Leadership	.15	-.22	-.33**	—		
Physical Condition	.22	.04	.10	.17	—	
Work Orientation	.16	.17	.10	.28*	.41**	—

N = 76. ** Correlation is significant at the $p < .01$ level. * Correlation is significant at the $p < .05$ level.

Correlations between AABio Scales and AAMIM Scales

The correlations between the AABio scales and the AAMIM scales are presented in Table 9. Generally speaking, the correlations between the scales from the two instruments demonstrated very good convergent and divergent validities. That is, scales that should be correlated with each other were, and scales that should not be correlated with each other were not. For example, AAMIM Leadership correlated most highly with AABio Peer Leadership ($r = .70, p < .01$) and AABio Diplomacy ($r = .50, p < .01$). AABio Fitness Motivation correlated significantly with AAMIM Leadership ($r = .25, p < .05$), AAMIM Physical Condition ($r = .53, p < .01$), and AAMIM Work Orientation ($r = .35, p < .01$), but did not correlate significantly with AAMIM Adjustment, AAMIM Agreeableness, or AAMIM Dependability. AAMIM Work Orientation correlated significantly with AABio Attention to Detail ($r = .34, p < .01$), AABio Fitness Motivation ($r = .35, p < .01$), AABio Multi-Tasking ($r = .33$, p $< .01$), and AABio Work Motivation ($r = .35, p < .01$). AABio Stress Tolerance correlated significantly with AAMIM Adjustment ($r = .50, p < .01$), AAMIM Leadership ($r = .32, p < .01$), and AAMIM Physical Condition ($r = .24, p < .05$). Finally, the AABio Unlikely Virtues scale correlated with the AAMIM Unlikely Virtues scale ($r = .42, p < .01$).

Army Aviation Information

The AAInfo measure included 50 multiple-choice questions designed to measure eight content areas (e.g., basic flight rules/knowledge). Correct answers were assigned a score value of one and incorrect answers were assigned a score value of zero. Scores for content areas and the total test were calculated by summing the appropriate item scores.

For this measure, means, standard deviations, and internal consistency reliabilities were calculated for each content area as well as the overall test. Item difficulties and item-total correlations were also calculated.

Table 9

Correlations between AABio and AAMIM Scales

	AAMIM Scales					
AABio Scales	Adjustment	Agreeableness	Dependability	Leadership	Physical Condition	Work Orientation
Adaptability	.31**	.01	-.25*	.31**	.02	.02
Army Aviation Identification	.15	-.06	.04	.17	-.03	.18
Attention to Detail	.28*	.21	-.05	.11	.21	.34**
Attitude Toward Authority	.24*	.24*	.29*	.09	.10	-.07
Cognitive Flexibility	.25*	-.04	-.22	.39**	-.04	-.01
Decisiveness	.09	.07	-.05	.26*	-.19	.00
Diplomacy	.22	.07	-.20	.50**	.07	.11
Fitness Motivation	.09	-.05	-.07	.25*	.53**	.35**
Internal Control	.33**	.06	-.06	.08	.02	.01
Multi-Tasking	.27*	.09	.07	.19	.25*	.33**
Peer Leadership	.22	-.11	-.21	.70**	.18	.15
Reasonable Risk-Taking	.09	.34**	.32**	-.14	-.03	.12
Risk Tolerance	.12	-.24*	-.24*	.29*	.06	.08
Stress Tolerance	.50**	.10	.02	.32**	.24*	.10
Work Motivation	.11	.02	.13	.27*	.18	.35**

N = 76. ** Correlation is significant at the $p < .01$ level. * Correlation is significant at the $p < .05$ level.

The results by total score and separately by subscore are presented in Table 10. On average, respondents answered 65% of the AAInfo items correctly; the average item-total correlation across all items was .25, and the alpha across all 50 items was .80. Some interesting patterns were found in examining the AAInfo subscores. The respondents performed most poorly on the questions dealing with meteorological conditions impacting helicopter flight, the conditions faced by helicopter pilots, and basic flight rules/knowledge. They performed best on questions dealing with types of Army helicopters, information specific to Army aviation, and the

basic operation of helicopters. In general, the vast majority of the AAInfo items demonstrated acceptable levels of item-total correlations, with the exception of those questions dealing with types of Army helicopters and information specific to Army aviation. The low item-total correlations for questions from these subscores, however, are clearly due to a lack of variance in performance on these items (i.e., the vast majority of respondents answered these items correctly). These items were kept in the pool because it was thought that applicants might not be as familiar with these subjects as the individuals included in the pilot test (currently enrolled trainees who were tested just days prior to beginning IERW).

As a result of these analyses, two items were dropped from the AAInfo. One item was dropped because very few respondents correctly answered the item (from basic flight rules/knowledge) and one item because it had a negative item-total correlation (from conditions faced by helicopter pilots).

Table 10

Descriptive Statistics of AAInfo Subscores

Scale	# Items	Proportion Correct	Average Item-total r
Basic flight rules/knowledge	5	.42	.22
Basic operation of helicopters	12	.74	.21
Conditions faced by helicopter pilots	5	.47	.27
Information specific to Army aviation	3	.83	.04
Major helicopter controls and parts	12	.66	.33
Meteorological conditions impacting helicopter flight	3	.40	.32
Physical forces impacting helicopter flight	7	.64	.31
Types of Army helicopters	3	.96	.10
Overall	50	.65	.25

$N = 80$.

Respondents were asked five questions designed to assess the extent to which prior knowledge and experience with aircraft might impact one's performance on AAInfo (Appendix C).

Analysis of variance (ANOVA) was then used to examine the extent to which responses to each of these questions impacted the respondents' AAInfo total score. Significant F statistics were found for all of the above questions except for the questions regarding frequency of playing flight simulation games and the number of times participants had flown in a helicopter. The results of the ANOVAs for the remaining questions are presented in Table 11.

Table 11

AAInfo Total Score Mean Differences by Amount of Knowledge and Experience with Flying and/or Army Aviation

	Proportion Correct	*SD*
How much did you know about helicopters and Army aviation before you applied to become an Army aviator? ($F = 8.05$, $p < .001$)		
Much more than others	.76 [a]	.13
More than others	.65 [b]	.10
About the same as others	.58 [c]	.10
Less than others	.59 [c]	.10
Much less than others	.50 [d]	.14
How much time did you spend learning about helicopters before you applied to become an Army aviator? ($F = 11.00$, $p < .001$)		
No time at all	.54 [a]	.11
1-4 hours	.61 [a]	.13
5-12 hours	.60 [a]	.08
13-24 hours	.59 [a]	.07
More than 24 hours	.75 [b]	.11
Do you have a private license to fly a helicopter or a fixed wing aircraft? ($F = 12.47$, $p < .001$)		
Yes, for both	.81 [a]	.13
Yes, helicopter only	—	—
Yes, fixed-wing aircraft only	.72 [a]	.14
No, but I have taken flying lessons	.74 [a]	.07
No	.59 [b]	.10

Note. Means with different superscripts within the results for each question are significantly different from each other.

The amount of knowledge that the respondents had with respect to helicopters and Army aviation impacted how well they performed on the AAInfo. Not surprisingly, the more knowledge that the participants possessed, the better they performed on the test. In addition, the amount of time they spent learning about helicopters prior to applying to become an Army aviator also impacted their total score. Specifically, those who spent more than 24 hours studying performed significantly better than those who spent less than 24 hours studying. Finally, those respondents who possessed a pilot's license or had taken flying lessons performed significantly better on the AAInfo than those who had no flying experience. No participants reported that they had a license to fly only helicopters. Thus, not unexpectedly, the amount of flight experience

and time respondents spent studying did have a significant impact on how well they performed on the AAInfo. It is clear, however, that a substantial amount of study is required before the candidates realize any benefits from such efforts.

Perceptual Speed and Accuracy

There were three PSA measures: Hidden Figures, Simple Drawings, and Panel Displays. The Hidden Figures measure included 30 items. The Simple Drawings measure included 100 items. The Panel Displays measure included 40 items. These measures were designed to be highly speeded, such that most participants could not complete all of the items in the measures. Correct answers to items on these measures were assigned a score value of one and incorrect answers or missing responses were assigned a score of zero. Participants were penalized for guessing and the total score for each measure was calculated as the total number of right answers minus 1/5 of the total number of wrong answers (there were five response options for these items).

The means, standard deviations, and split-half reliabilities were calculated for each test. The split-half reliabilities were calculated by creating total scores for odd and even items, correlating these two scores, and conducting a Spearman-Brown correction on the correlation for each measure. The item-total correlations were also examined within each measure. Finally, the intercorrelations among the three PSA measures were calculated.

The descriptive statistics for the PSA tests are presented in Table 12. For the Hidden Figures test, the participants answered an average of 16.71 items correctly and 10.08 items incorrectly. The test exhibited a very high split-half reliability (Spearman-Brown correction applied) of .90. Finally, 58% of the sample finished all of the items, so it was clear that either the time limit needed to be shortened or more items added. It was decided to keep the time limit at 6 minutes, and 20 items were added for use during the preliminary validation, to follow.

For the Simple Drawings test, the participants answered an average of 44.29 items correctly and 1.19 items incorrectly. Again, the Simple Drawings test exhibited a very high split-half reliability (Spearman-Brown correction applied) of .93. Approximately 13% of the sample finished 53 items and only 2.5% of the sample finished 61 items, so the number of items on this test could have been decreased. Given the difficulty of doing so on the current testing platform, however, it was decided to leave the test length as is, and increase the time limit from 2 minutes to 3 minutes for the preliminary validation research.

Table 12

Descriptive Statistics of PSA Tests

Scale	Mean	*SD*
Hidden Figures (30 items)		
Number correct	16.71	7.13
Number incorrect	10.08	7.22
Score corrected for guessing	14.70	8.31
Simple Drawings (100 items)		
Number correct	44.29	6.07
Number incorrect	1.19	1.17
Score corrected for guessing	44.05	6.04
Panel Displays (40 items)		
Number correct	21.29	4.75
Number incorrect	1.68	1.64
Score corrected for guessing	20.95	4.84

N = 80.

For the Panel Displays test, the participants answered an average of 21.29 items correctly and 1.68 items incorrectly. Approximately 48% of the sample finished 24 items, 15% of the sample finished 29 items, and less than 1% of the sample finished 34 items. So, once again, it is clear that there are more items on this test than are needed with a 2-minute time limit. However, instead of deleting items or increasing the time limit, it was decided to drop this test entirely. It was moderately correlated (see Table 13) with the Simple Drawings test ($r = .47, p < .01$), so that it may not add much predictive validity if used in the validation research to follow. It also correlated .27 ($p < .05$) with AAInfo. Thus, there is some indication that, because the stimuli included in the Panel Displays test contains pictures of actual aircraft gauges, the participants who possess knowledge regarding such gauges may be at an advantage over those who do not.

Table 13

Correlations between PSA Measures and AAInfo

	Hidden Figures	Simple Drawings	Panel Displays
Simple Drawings	.37**	—	
Panel Displays	.25*	.47**	—
AAInfo	.16	.00	.27*

N = 80. ** Correlation is significant at the $p < .01$ level. * Correlation is significant at the $p < .05$ level.

Popcorn Test

For the Popcorn Test means and standard deviations were calculated for all 10 trials and for several subsets of trials (e.g., 8^{th}, 9^{th}, and 10^{th} trials), separately for the two data collection periods as well as for the entire group of participants. Recall that each Popcorn trial score represents the percentage of total possible points that the respondent received (i.e., the ratio of points attained/points possible). The descriptive statistics for the entire group of participants, by trial, and then aggregated separately across all 10 trials and again across just the last three trials, are presented in Table 14.

Recall that the pilot test data collection occurred over two periods. During the first testing period technical problems resulted in data from two trials being lost (trials 4 and 8). Thus, for these two trials the sample sizes are approximately half that of the sample sizes for the other eight trials. The learning curve from the first round of data collection showed non-monotonic improvement in performance. More importantly, the intertrial matrix did not show evidence of a superdiagonal form. Again, these problems were attributed to between-trial differences in the total number of targets presented and their average speed. The two trials that differed the most in number of targets and the average target speed were replaced with two of the existing trials for the second data collection. Thus, the second testing session had two pairs of identical trials, that is, only eight of the 10 trials were unique.

The second data collection showed the best results. The intertrial correlation matrix showed essentially perfect superdiagonal form. However, the test did not reach differential stability, apparently due to lack of practice. Therefore, five additional trials were added to the program for a total of 15 trials. Again, these additional trials were repetitions of existing trials. No trial occurred more than twice in the sequence and each repetition was separated by at least four trials from its parent trial. The final version of Popcorn was pretested by volunteers.

Table 14

Descriptive Statistics for Popcorn

	Trial										All Trials	Last 3 Trials
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**		
N	78	78	78	43	78	78	78	43	78	78	78	78
Mean	.45	.45	.54	.51	.48	.52	.57	.53	.53	.59	.51	.54
SD	.11	.10	.10	.10	.09	.09	.10	.09	.10	.07	.08	.07

Respondents were asked five questions in order to assess the extent to which experience with computer technology and playing video games might impact one's performance on the Popcorn test (Appendix C).

ANOVAs were then conducted using three Popcorn scores: the average across all 10 trials, the average across the last 3 trials, and the score on the 10th trial only. Significant F statistics were found for the amount of experience with non-flight simulation video game experience, and to a lesser degree with amount of experience with flight simulation video games. The results of the ANOVA for these two questions are shown in Table 15.

In terms of experience with non-flight simulation video games, post hoc t-tests showed that respondents who had never played a non-flight simulation video game performed significantly more poorly on Popcorn, across all three measures, than did those respondents who had any experience with non-flight simulation video games. The impact of experience with flight simulation video games was less pronounced. The only significant F statistic found for amount of experience with flight simulation video games was for the score on the 10th Popcorn trial. Again, those respondents who had no experience whatsoever with flight simulation video games performed more poorly on the 10th trial of Popcorn than did those respondents who had at least some experience with flight simulation video games.

Table 15

Popcorn Mean Differences by Amount of Experience with Flight Simulation and Non-Flight Simulation Video Games

	Almost everyday		A few days each week		A few days each month		Less than once/ month		Never		F	Sig. Level
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
In the past few years, how often have you played video games other than flight simulation games?												
All 10 Trials	.57 [a]	.09	.52 [a]	.07	.52 [a]	.07	.49 [a]	.09	.45 [b]	.09	3.97	.006
Trials 8-10	.60 [a]	.08	.55 [a]	.06	.54 [a]	.07	.53 [a]	.09	.45 [b]	.12	3.88	.007
10th Trial Only	.63 [a]	.08	.60 [a]	.07	.60 [a]	.07	.58 [a]	.08	.50 [b]	.15	5.36	.001
In the past few years, how often have you played flight simulation games?												
All 10 Trials	.53	.06	.48	.08	.52	.09	.52	.06	.46	.11	2.01	.10
Trials 8-10	.56	.09	.54	.09	.54	.08	.55	.07	.48	.14	1.96	.11
10th Trial Only	.64 [a]	.05	.57 [a]	.06	.60 [a]	.09	.60 [a]	.06	.51 [b]	.16	3.13	.02

Note. Means with different superscripts within a row are significantly different from each other.

Correlations between Popcorn and PSA Tests

The correlations between the Popcorn measures and the PSA tests were also computed. These correlations are reported in Table 16. The average across all 10 Popcorn trials correlated moderately with scores on the Hidden Figures Test, and the average across the last three trials of Popcorn (8-10) correlated moderately with scores on the Simple Drawings Test. No other correlations were statistically significant.

Table 16

Selected Correlations between Popcorn and PSA Tests

	Hidden Figures	Simple Drawings	Panel Displays
Popcorn Trials 8-10	.18	.27*	.22
Popcorn 10th Trial Only	.15	.22	.16
Popcorn All 10 Trials	.31**	.21	.17

$N = 78$. ** Correlation is significant at the $p < .01$ level. * Correlation is significant at the $p < .05$ level.

Feedback Regarding Test Instructions and Utility of Test Battery

The pilot test participants were asked to provide feedback regarding the clarity of the instructions for the test battery, as well as how useful they thought this test battery would be for selecting helicopter pilots (Table 17). The majority (95%) said that the test instructions were either "Somewhat clear" (24%) or "Very clear" (71%). In addition, (although the participants were minimally familiar with the job of helicopter pilot, since they had not yet begun their training), 50% of the participants agreed or strongly agreed that the test battery measured skills important for becoming a helicopter pilot, whereas only 7.5% of the participants disagreed with this statement. The remaining participants neither agreed nor disagreed with this statement.

Table 17

Feedback Regarding Test Instructions and Utility of Test Battery

	Frequency	Percent
In general, how clear were the test instructions?		
Very clear	57	71.2
Somewhat clear	19	23.8
Somewhat unclear	2	2.5
Very Unclear	2	2.5
This test battery measures skills important for becoming a helicopter pilot.		
Strongly Agree	5	6.2
Agree	35	43.8
Neither agree nor disagree	34	42.5
Disagree	2	2.5
Strongly disagree	4	5.0

$N = 80$.

Preliminary Validation Version of Predictor Battery

Table 18 presents a summary of the revisions made to the non-AACog tests selected to be included in the predictor battery during the preliminary validation effort to follow. Recall that the AACog tests are from the Navy's ASTB, so they can only be administered as they currently exist during the preliminary validation study.

Table 18

Summary of Revisions

Test	Original # of Items	Revised # of Items
Army Aviation Information	50	48
Hidden Figures	30	50
Simple Drawings	100	100
Panel Displays	40	0
Popcorn Test	10	15
Army Aviation Biodata	154	154

Finally, Table 19 shows the predictors chosen to be administered in the preliminary validation research and the estimated time it takes to complete them. The entire battery takes approximately 6 hours to complete (not including lunch), and the order of administration of these tests was counterbalanced in the validation research. The methods, analyses, and results of the SIFT validation research will be described in a future report.

Table 19

Predictors for Preliminary Validation Research

Test	Administration Time
Army Aviation Cognitive Test (AACog)	150 minutes
Army Aviation Measure of Motivation (AAMIM)	30 minutes
Popcorn Test	45 minutes
Perceptual Speed and Accuracy Tests (PSA)	15 minutes
Army Aviation Information (AAInfo)	25 minutes
Army Aviation Biodata (AABio)	45 minutes

Conclusions

The development of SIFT followed a systematic process, from a thorough assessment of training and job requirements to the pilot testing of a prototype battery, described herein. The job analysis, in conjunction with the results of a focused pilot selection literature review, led to the selection of the following predictor measures for inclusion in a prototype battery for pilot testing:

- *Cognitive ability*: Including all cognitive subtests from the Navy's Aviator Selection Test Battery (ASTB)
- *Perceptual Speed & Accuracy*: Using a newly-developed test, specifically designed for Army aviation selection
- *Personality/Temperament*: Using the Army Assessment of Individual Motivation (AIM) and the Test of Adaptive Personality (TAP)
- *Motivation/Attitude*: Using a newly-developed Army Aviation Information Test and the Army Aviation Identification Scale
- *Task Prioritization*: Using the computer-based "Popcorn Test"

The results of the pilot test of these measures resulted in revisions and decisions as to the predictors to be included in the prototype battery for preliminary validation, which was the next step in this development process. The final steps in this effort will be to use this same systematic process to develop a classification instrument for Army aviation and a selection instrument for Unmanned Aviation Systems (UAS). A computer-based, web-administered instrument will be developed to assess the relevant attributes of applicants for UAS operator training. The same methodology described in this report will be used to produce a scientifically sound instrument to predict the likelihood that individuals will successfully complete training to perform as UAS operators. Regarding the classification instrument, the objectives will be to develop a computer-based battery to determine the differential suitability of aviation students to the various Army aircraft, and to develop an automated algorithm to assign students to training tracks while they are still in initial training.

References

Christal, R. L. (1975). Personality factors in selection and flight proficiency. *Aviation, Space and Environmental Medicine, 46,* 309-311.

Daryanian, B. (1980). *Subjective scaling of mental workload in a multi-task environment.* Paper presented at the Sixteenth Annual Conference on Manual Control. Cambridge, MA: Massachusetts Institute of Technology.

Grice, R., & Katz, L. C. (2006). *Personality Profiles of Experienced U.S. Army Aviators Across Mission Platforms* (ARI Technical Report 1185). Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Hart, S. G., Battiste, V., & Lester, P. T. (1984). POPCORN: A supervisory control simulation for workload and performance research. *20th Annual Conference on Manual Control* (Vol. VI, pp. 431-453).

Held, J. D., & Farmer, W. L. (2004). *Knowledge and skills tests as measures of motivation and interest.* (NPRDC Letter report Ser 3900, 13/000056, 9 Sep 2004). Millington, TN: Navy Personnel Research, Studies and Technology.

Kilcullen, R. N. (2004). *Test of Adaptable Personality.* Presentation given at the ARI Special Forces User Review, Arlington, VA, November 17, 2004.

Kubisiak, C., & Katz, L. C. (2006). *U.S. Army Aviator Job Analysis* (ARI Technical Report 1189). Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Melton, A. W. (1947). *Apparatus Tests Report No. 4.* Washington, D.C.: U.S. Government Printing Office.

Pattipati, K. R., Kleinman, D. L., & Ephrath, A. R. (1983). A dynamic decision model of human task selection performance. *IEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 145-166.

Paullin, C., Bruskiewicz, K.T., Houston, J., & Damos, D. (2005). *Development of a Selection Instrument for Army Flight Training (SIFT): Review of the literature and identification of a selection strategy* (Institute Report #493). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Paullin, C., Katz, L. C., Bruskiewicz, K. T., Houston, J., & Damos, D. (2006). *Review of Aviator Selection* (ARI Technical Report 1183). Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Retzlaff, P. D., King, R. E., McGlohn, S. E., & Callister, J. D. (1996). *The development of the Armstrong Laboratory Aviation Personality Survey (ALAPS)* (AL/AO-TR-1996-0108). Brooks AFB, TX: Armstrong Laboratory, Air Force Materiel Command.

Roscoe, S. N. (1980). *Aviation psychology.* Ames, Iowa: The Iowa State University.

Roscoe, S. N., Corl, L., & LaRoche, J. (1997). *Predicting human performance.* Pierrefonds, Quebec: Helio Press, Inc.

Russell, R. L., Peterson, N. G., Rosse, R. L., Hatten, J. T., McHenry, J. J., & Houston, J. S. (2001). The measurement of cognitive, perceptual, and psychomotor abilities. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 64-89). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Tulga, M. K., & Sheridan. T. B. (1980). Dynamic decisions and work load in multitask supervisor control. *IEE Transactions on Systems, Man, and Cybernetics, SMC-10,* 217-231.

White, L. A., & Young, M. C. (1998). *Development and validation of the Assessment of Individual Motivation (AIM).* Paper presented at the Annual Meeting of the American Psychological Association, San Francisco, CA.

Appendix A

Classroom XXI Computer Specifications

Monitor Specifications: 17" Flat Screen Monitors (Sony Trinitron Multiscan 9-220R).
Monitor Resolution: 1024 X 768 DPI.
Computer RAM: 256 K.
Hard Drive Processing Speed: PIII 933MHZ.
Operating System: Windows 2000, Version 5.00.2195 W/SP4.
Internet Browser: Internet Explorer 6.0
Internet Connection Speed: 100.0 Mbps, connected through Ft. Rucker's T-2 server using a 10/100 ethernet card.
ATI Video Card: RAGE 128 GL AGP.

Appendix B

Popcorn Debriefing Questionnaire

1. To get the highest score possible on the test you just finished, what type of targets do you have to erase?
   __ Large targets, no matter how fast they are moving
   __ Fast moving targets, no matter what their size is
   __ Large targets that are moving fast
   __ Small targets that are moving fast
   __ None of the above are correct

2. Did you develop a strategy for the test you just finished?
   __ Yes [***please answer question 3***]
   __ No

3. [***Answer if you responded "Yes" to Question 2***] Please describe the strategy you used briefly in the space provided below.

Appendix C

Computer and Video Game Experience Questionnaire

- Compared to others your age, how much computer knowledge do you have?
  - Much more than others
  - More than others
  - About the same as others
  - Less than others
  - Much less than others
- How much experience do you have using computers?
  - Extensive experience
  - Great deal of experience
  - Moderate amount
  - Small amount
- How much experience do you have using a computer mouse?
  - Extensive experience
  - Great deal of experience
  - Moderate amount
  - Small amount
- In the past few years, how often have you played flight simulation games?
  - Almost everyday (25-30 days per month)
  - A few days each week (8-24 days per month)
  - A few days each month (1-7 days each month)
  - Less often than once per month
  - I have never played a flight simulation video game
- In the past few years, how often have you played video games other than flight simulation games?
  - Almost everyday (25-30 days per month)
  - A few days each week (8-24 days per month)
  - A few days each month (1-7 days each month)
  - Less often than once per month
  - I have never played a non-flight simulation video game